MODUL 11

INDUKSI DAN ATURAN ASOSIASI

11.1. Tujuan

1. Mahasiswa mampu menggunakan induksi aturan, dan aturan asosiasi.
2. Mahasiswa mampu menerapkan aturan induksi aturan, dan aturan asosiasi dalam kasus nyata.

11.2. Landasan Teori

Rule induction adalah salah satu teknik dalam data mining yang paling sering digunakan untuk menemukan pengetahuan dalam sistem *unsupervised learning*. *Rule* (aturan) adalah bentuk sederhana dari "jika ini maka ini dan ini dan kemudian ini". Sebagai contoh: jika seseorang membeli roti maka orang tersebut juga cenderung untuk membeli selai.

Agar aturan-aturan tersebut bermanfaat maka harus ditambahkan dua informasi tambahan sesuai dengan keadaaan sebenarnya yaitu :

1. Keakuratan (*accuracy/confidence*) yang menunjukkan seberapa sering aturan tersebut benar.
2. Penerapan (*coverage/support*) yaitu angka yang menunjukkan seberapa sering aturan tersebut dipakai.

Association Rule merupakan suatu proses untuk menemukan semua aturan *assosiatif* yang memenuhi syarat *minimum* untuk *support* (*minsup*) dan syarat *minimum* untuk *confidence* (*minconf*) pada sebuah *database.*

©Muhammadiyah university Press

Dalam menentukan suatu *Association Rule* umumnya terdapat dua ukuran kepercayaan (*interestingness measure*), yaitu *support* dan *confidence*. Kedua ukuran ini akan digunakan untuk *interesting association rules* dibandingkan dengan batasan yang telah ditentukan. Batasan inilah yang terdiri dari *minsup* dan *minconf*. *Assosiation Rule Mining* adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *dataset*. Dimulai dengan mencari *frequent itemset*, yaitu kombinasi yang paling sering terjadi dalam suatu itemset dan harus memenuhi *minimum support*.

Dalam tahap ini akan dicari kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Untuk mendapatkan nilai *support* untuk sebuah item A dapat diperoleh dari rumus berikut :

$$Support(A) = \frac{Jumlah\ transaksi\ yang\ mengandung\ item\ A}{Total\ transaksi}$$

Sementara itu, untuk mencari nilai *support* dari 2-*item* dapat diperoleh dari rumus berikut :

$$Support\ (A, B) = P(A \cap B)$$

$$P(A \cap B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Total\ transaksi}$$

Setelah semua *frequent item* dan *Large itemset* ditemukan, dapat dicari semua *Association Rules* yang memenuhi syarat *minimum* untuk *confidence* (*minconf*) dengan menggunakan rumus berikut ini:

$$Confidence\ (A \rightarrow B) = P(B|A)$$

$$P(B|A) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ yang\ mengandung\ item\ A}$$

©Muhammadiyah University Press

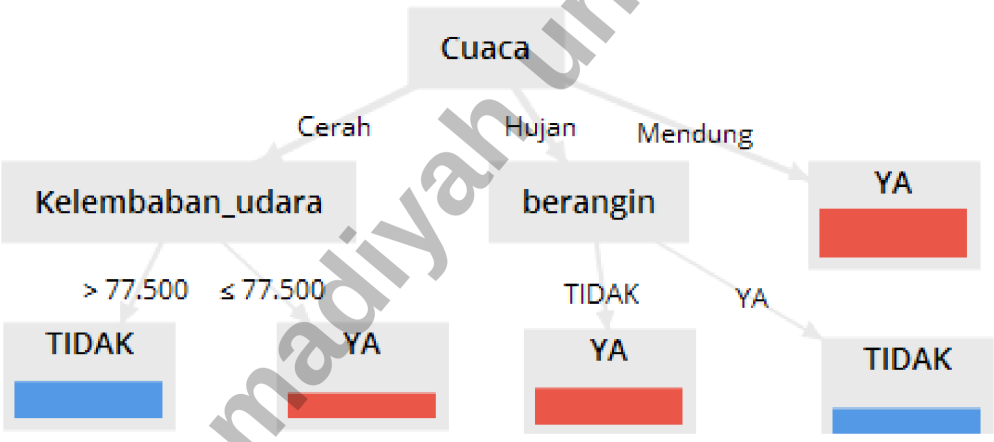
11.3. Alat dan Bahan

1. Komputer dengan sistem operasi Windows.
2. Program aplikasi RapidMiner.
3. Modul Praktikum Data Warehousing dan Data Mining.

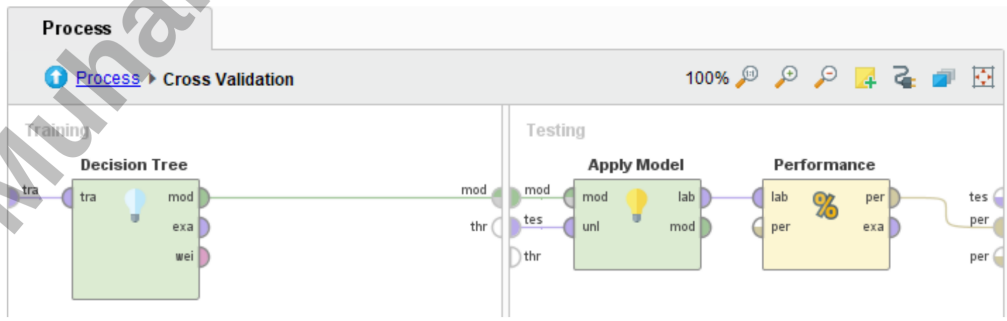
11.4. Langkah-langkah Praktikum

11.4.1. Induksi Aturan Data Cuaca

1. Jalankan aplikasi RapidMiner.
2. Gunakan model proses dari praktikum data mining **Modul 9 Kegiatan 9.4.2**. (Jika tidak disimpan, lakukan kembali kegiatan tersebut).
3. Dari hasil kegiatan tersebut menghasilkan sebuah pohon keputusan seperti gambar berikut.

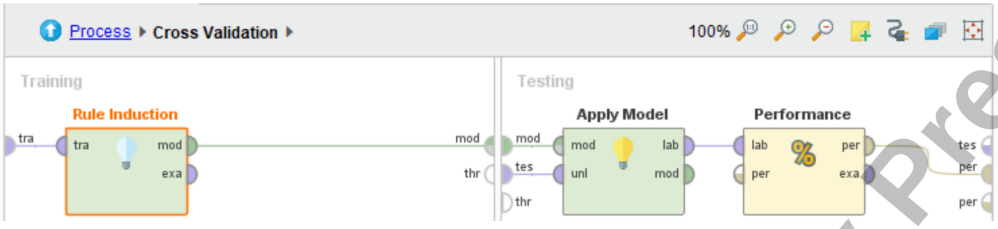


4. Kembali ke **Process – Cross Validation**

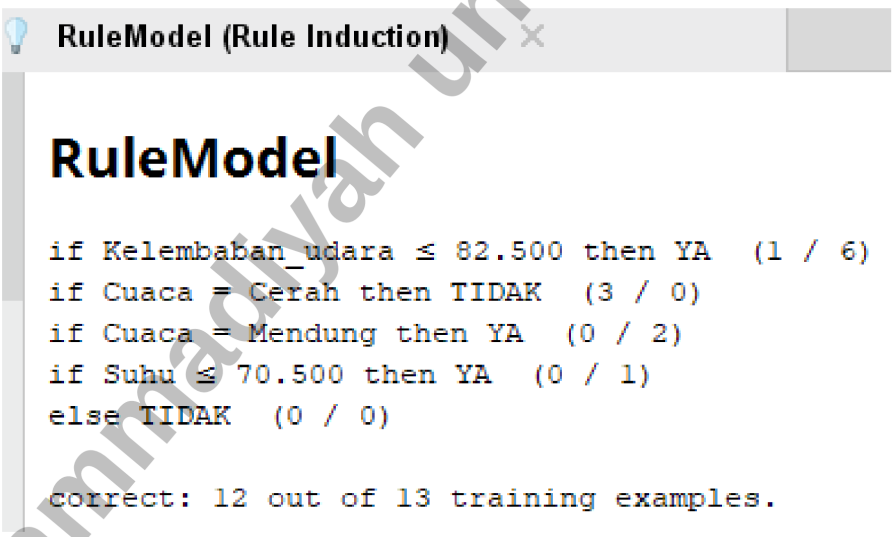


©Muhammadiyah University Press

5. Klik kanan operator **Decision Tree** pada area **Training**.
6. Pilih **Replace Operator → Modeling → Predictive → Rules → Rule Induction**.



7. Tanpa mengubah parameter-parameter yang lainnya, jalankan proses dengan menekan tombol **Run** (atau menekan tombol F11).
8. Sehingga akan diperoleh sebuah induksi aturan dari data training yang diberikan yang disebut sebagai **RuleModel (Rule Induction)**.



9. Dari hasil RuleModel tersebut dapat dilihat hubungan sebab akibat (Implikasi dengan menggunakan pernyataan "jika... maka..." atau "if... then...") dari parameter-parameter yang diberikan. Aturan tersebut bisa digunakan sebagai dasar untuk pengambilan keputusan apakah seseorang akan bermain tenis atau tidak berdasarkan data cuaca yang ada.

©Muhammadiyah university Press

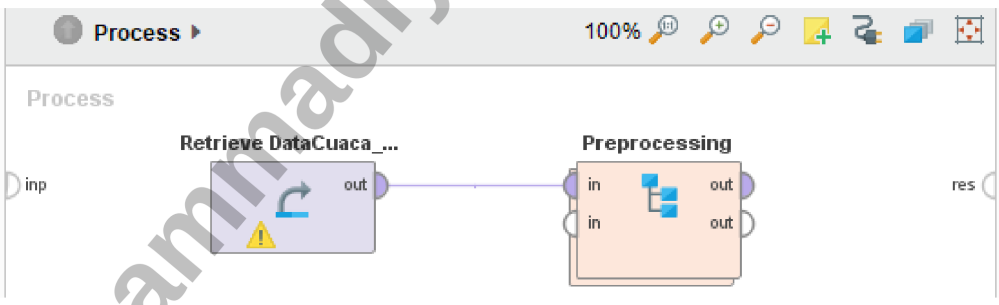
10. Parameter *correct* menunjukkan jumlah data benar dalam Induksi Aturan terhadap total data yang digunakan sebagai data training.
11. Model Rule Induction ini juga bisa ditunjukkan hasil Performance Vector (Performance).

accuracy: 65.00% +/- 45.00% (micro average: 71.43%)

	true TIDAK	true YA	class precision
pred. TIDAK	2	1	66.67%
pred. YA	3	8	72.73%
class recall	40.00%	88.89%	

11.4.2. Aturan Asosiasi Data Cuaca

1. Masih dengan menggunakan RapidMiner, pilih **New Process**.
2. Gunakan **DataCuaca_Training** dan drag dari *repository* ke area **Process View**.
3. Tambahkan operator **Utility → Subprocess** ke dalam area. Ubah nama operator ini menjadi **Preprocessing** dengan klik kanan operator **Subprocess – Rename**. Hubungkan port output Retrieve dengan port input Preprocessing.

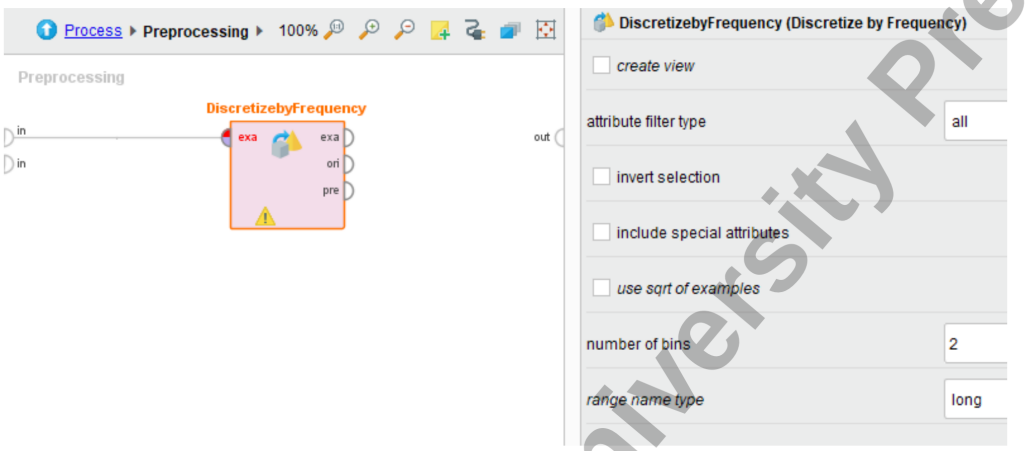


4. Klik ganda operator **Preprocessing** sehingga masuk pada area *Nested Chain*.
5. Pada area *Nested Chain* ini tambahkan operator-operator berikut:
   a) **Cleansing → Binning → Discretize by Frequency**

      Ubah nama operator ini menjadi

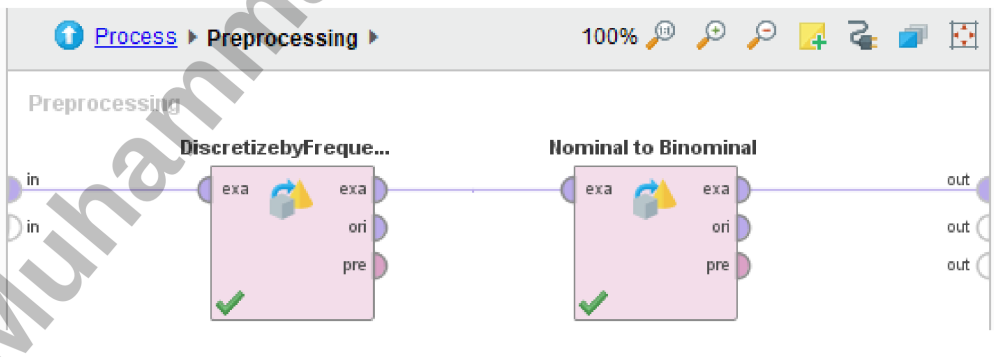
©Muhammadiyah University Press

"**DiscretizebyFrequency**" dan biarkan nilai parameter **number of bins** (jumlah interval) = 2 (anda juga bisa mengubah-ubah nilai ini, misal = 5). Hubungkan panel **in** pada masukan example set dengan port input examination operator ini.



b) **Blending → Attributes → Types → Nominal to Binomial**

Ubah nama operator ini menjadi "**Nominal2Binominal**", hubungkan output operator **DiscretizebyFrequency** dengan masukan examination pada operator ini, dan output examination operator ini dengan panel out.



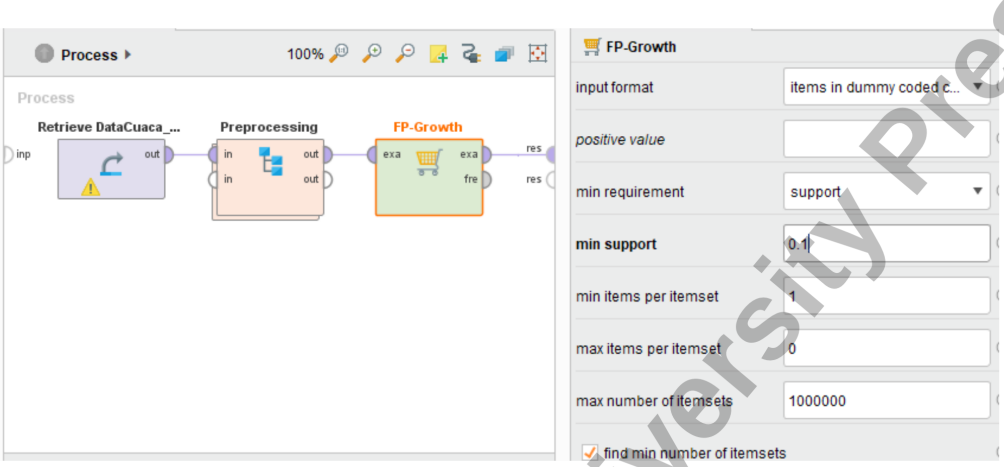
6. Kembali ke *main process*, tambahkan 2 buah operator:

a) **Modeling → Associations → FP-Growth**

Ubah nilai parameter pada min support = 0.1,

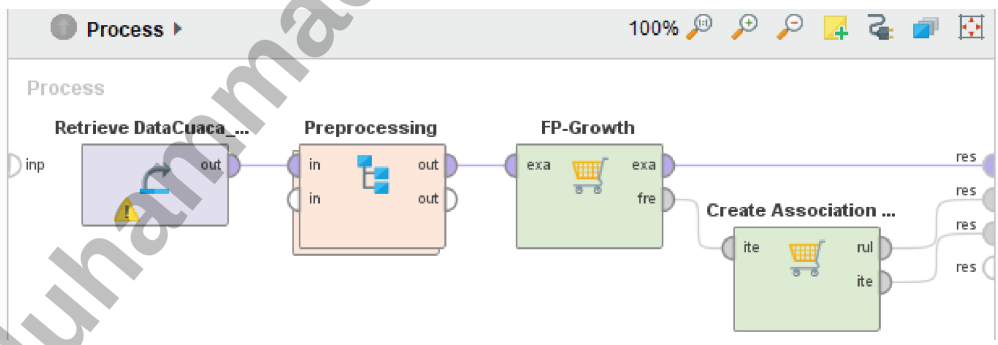
©Muhammadiyah University Press

hubungkan port out operator Preprocessing dengan port entry exa pada operator ini dan output exa dengan connector res (result, resultat) pada work area.



b) Modeling → Associations → Create Association Rules

Hubungkan output fre (frequent sets, frequent sets) pada operator FP-Growth dengan masukan ite (item sets, itemsets) pada operator ini, output rul (rules) dengan connector res dalam work area, dan ouput ite (item sets) pada operator ini dengan connector res lainnya dalam work area.



7. Jalankan proses dengan menekan tombol **Run** (atau menekan tombol F11).
8. Dapat dilihat hasil-hasil aturan asosiasi sebagai berikut:

©Muhammadiyah University Press

a) Frequent Item Set (FP-Growth)

No. of Sets: 26
Total Max. Size: 4

Min. Size: 1
Max. Size: 4
Contains Item:
Update View

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.500	Kelembaban_udara			
1	0.429	Suhu			
1	0.429	berangin			
1	0.357	Cuaca = Cerah			
1	0.357	Cuaca = Hujan			
1	0.286	Cuaca = Mendung			
2	0.214	Kelembaban_udara	Suhu		
2	0.214	Kelembaban_udara	berangin		
2	0.214	Kelembaban_udara	Cuaca = Cerah		
2	0.143	Kelembaban_udara	Cuaca = Hujan		
2	0.143	Kelembaban_udara	Cuaca = Mendung		
2	0.143	Suhu	berangin		
2	0.214	Suhu	Cuaca = Cerah		
2	0.071	Suhu	Cuaca = Hujan		
2	0.143	Suhu	Cuaca = Mendung		

Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 26 set, dan jumlah total max size = 4, yang terdiri dari 4 buah itemset. Kita bisa menyortir berdasarkan kolom dengan mengklik header kolom. Dengan menekan tombol CTRL selama mengklik, membuat kita dapat memilih beberapa kolom untuk disortir.

b) Association Rules (Create Association Rules)

i. Table View

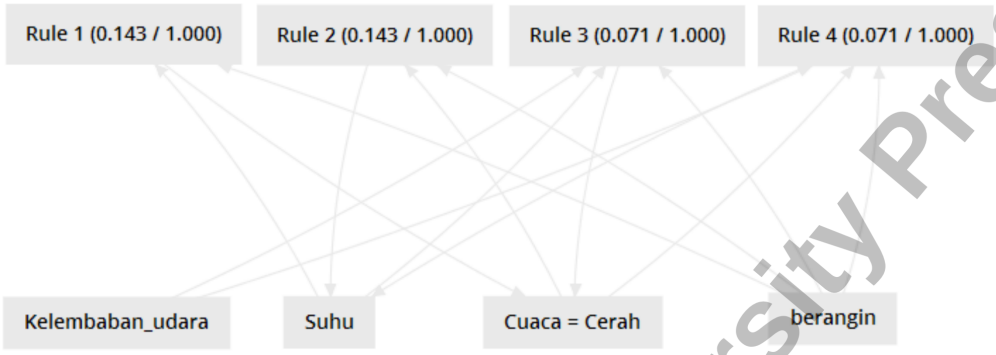
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
1	Suhu, berangin	Cuaca = Cerah	0.143	1	1	-0.143	0.092
2	berangin, Cuaca = Cerah	Suhu	0.143	1	1	-0.143	0.082
3	Kelembaban_udara, Suhu, berangin	Cuaca = Cerah	0.071	1	1	-0.071	0.046
4	Kelembaban_udara, berangin, Cuaca = Cerah	Suhu	0.071	1	1	-0.071	0.041

Pada tabel ini dapat dilihat terdapat 4 data pasangan premis-premis dan kesimpulan (Premises and Conclusion) yang menunjukkan hubungan implikasi berdasarkan aturan asosiasi. Aturan ini diperoleh hasil perhitungan asosiasi

©Muhammadiyah University Press

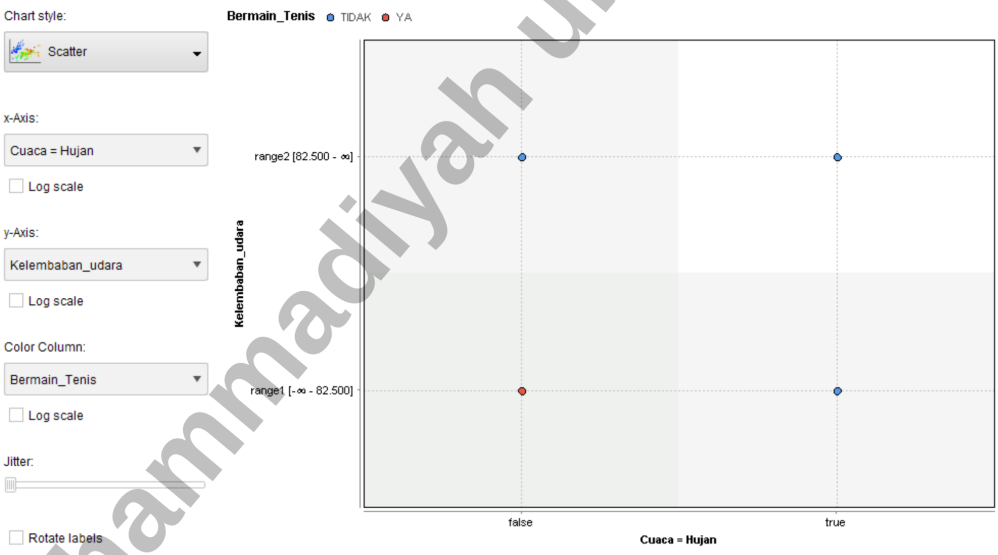
dengan nilai Support, Confidence, LaPlace, Gain, p-s, dan Lift pada masing-masing pernyataan.

ii. Graph View



Pada mode ini diilustrasikan pola hubungan sebab akibat dalam bentuk grafik.

c) ExampleSet (Nominal2Binomial) → Charts View



Pada hasil ini dapat ditunjukkan pola distribusi masing-masing data cuaca yang diproses menggunakan aturan asosiasi. Kita bisa mengubah-ubah variabel dalam x-Axis dan y-Axis serta nilai Jitter. Namun pastikan bahwa pada *Color Column* adalah variabel *dependent* yaitu Bermain_Tenis.

©Muhammadiyah university Press

11.5. Tugas

1. Dengan menggunakan data Lama Studi (format Excel) pada **Tugas Modul 6 Soal No. 1**, carilah pola hubungan berdasarkan *Induction Rule* (*Rule Model*), dan nilai *performance vector*!
2. Masih dengan menggunakan *data training* yang sama, dengan ketentuan bahwa pada operator *Discretize by Frequency* memiliki nilai:

   a) *number of bins* = 2

   b) *number of bins* = **gunakan nilai yang lain**

   Carilah masing-masing nilai berikut berdasarkan *number of bins*-nya:

   a) Jumlah set aturan asosiasi dan total max size yang terbentuk berdasarkan FP-Growth (table view)! Gambarkan tabelnya!

   b) Jumlah data pasangan premis dan kesimpulan pada Association Rules (Create Association Rules)! Gambarkan tabelnya! Gambarkan pula grafik yang terbentuk!

   c) Gambarkan grafik chart pola distribusi data pada ExampleSet yang terbentuk!

©Muhammadiyah university Press